

# Florida Weights for Wave 1, BYU-Wisconsin Election Panel

Charles H. Franklin  
*University of Wisconsin, Madison*

Wednesday, June 30, 2004 at 12:33

## 1 Florida Weighting in Wave 1

When we revised the sample design to make Florida an oversample stratum, the design I used assumed 400 cases for the Florida stratum. In wave 1, we are instead taking 201 cases there, 199 fewer than the oversample would provide. Our total wave 1  $n$  is now  $3001 - 199 = 2802$ . The 201 Florida cases are what was initially allocated to Florida in the national sample, before we replaced the Pennsylvania oversample with Florida.

This causes no substantial problems, but does mean that the weight for Florida needs to be adjusted for the smaller stratum sample. *Weights for other strata are ALSO affected by this change because we want the weighted  $n$  to sum to the actual number of respondents, 2802, rather than the original sample design of 3001.*

Because of the smaller Florida sample, the standard errors for Senate and Presidential battlegrounds, and for the national population, are slightly changed by this revision. These are quite modest increases.

The code below repeats our sample calculations as before, but now computes the standard errors and margin of error based on the Florida sample size of 201 for wave 1. The weights for all strata are also recalculated.

The summary of results, and the table of weights (Table 3), appears in the final section below.

```
> rm(list = ls(all = TRUE))
> setwd("c:/data/campaign2004/news1absurvey")
> library(foreign)
```

```

> library(xtable)
> pop <- read.dta("statetoandpop.dta")
> attach(pop)
> pop$bgpres[state == "CO"] <- 1
> pop$bgpres[state == "LA"] <- 1
> attach(pop)
> pop$bg <- bgpres
> pop$bg[bgpres == 0 & bgpres == 0] <- 0
> pop$bg[bgpres == 0 & bgpres == 1] <- 1
> pop$bg[bgpres == 1] <- 2
> pop$bg[state == "OH"] <- 3
> pop$bg[state == "FL"] <- 4
> attach(pop)
> pop3 <- subset(pop, state != "AK" & state != "HI")
> attach(pop3)
> table(bg)

```

```

bg
 0  1  2  3  4
22  8 17  1  1

```

```

> table(state, bg)

```

```

      bg
state 0 1 2 3 4
  AL 1 0 0 0 0
  AR 0 0 1 0 0
  AZ 0 0 1 0 0
  CA 0 1 0 0 0
  CO 0 0 1 0 0
  CT 1 0 0 0 0
  DC 1 0 0 0 0
  DE 1 0 0 0 0
  FL 0 0 0 0 1
  GA 0 1 0 0 0
  IA 0 0 1 0 0
  ID 1 0 0 0 0
  IL 0 1 0 0 0
  IN 1 0 0 0 0
  KS 1 0 0 0 0

```

```
KY 0 1 0 0 0
LA 0 0 1 0 0
MA 1 0 0 0 0
MD 1 0 0 0 0
ME 0 0 1 0 0
MI 0 0 1 0 0
MN 0 0 1 0 0
MO 0 0 1 0 0
MS 1 0 0 0 0
MT 1 0 0 0 0
NC 0 1 0 0 0
ND 1 0 0 0 0
NE 1 0 0 0 0
NH 0 0 1 0 0
NJ 1 0 0 0 0
NM 0 0 1 0 0
NV 0 0 1 0 0
NY 1 0 0 0 0
OH 0 0 0 1 0
OK 0 1 0 0 0
OR 0 0 1 0 0
PA 0 0 1 0 0
RI 1 0 0 0 0
SC 0 1 0 0 0
SD 0 1 0 0 0
TN 1 0 0 0 0
TX 1 0 0 0 0
UT 1 0 0 0 0
VA 1 0 0 0 0
VT 1 0 0 0 0
WA 0 0 1 0 0
WI 0 0 1 0 0
WV 0 0 1 0 0
WY 1 0 0 0 0
```

```
> Npop <- sum(to2000)
> Npoph0 <- sum(to2000[bg == 0])
> Npoph1 <- sum(to2000[bg == 1])
> Npoph2 <- sum(to2000[bg == 2])
> Npoph3 <- sum(to2000[bg == 3])
```

```

> Npoph4 <- sum(to2000[bg == 4])
> Npoph <- c(Npoph0, Npoph1, Npoph2, Npoph3, Npoph4)
> n <- 3000
> pq <- c(0.65 * 0.35, 0.25, 0.25, 0.25, 0.25)
> nhnps <- c(600, 500, 1100, 400, 400)
> whnps <- nhnps/((Npoph/sum(Npoph)) * n)
> pop3$ninps[bg == 0] <- whnps[1] * to2000[bg == 0]/Npop *
+   n
> pop3$ninps[bg == 1] <- whnps[2] * to2000[bg == 1]/Npop *
+   n
> pop3$ninps[bg == 2] <- whnps[3] * to2000[bg == 2]/Npop *
+   n
> pop3$ninps[bg == 3] <- whnps[4] * to2000[bg == 3]/Npop *
+   n
> pop3$ninps[bg == 4] <- whnps[5] * to2000[bg == 4]/Npop *
+   n
> attach(pop3)
> sum(ninps)

```

```
[1] 3000
```

```

> pop3$ninps <- round(pop3$ninps + 0.01, digits = 0)
> pop3$ninps[state == "FL"] <- 201
> attach(pop3)
> sum(ninps)

```

```
[1] 2802
```

```
> print(cbind(state, ninps, bg sen, bg pres, bg))
```

	state	ninps	bg sen	bg pres	bg
[1,]	"AL"	"29"	"0"	"0"	"0"
[2,]	"AZ"	"57"	"0"	"1"	"2"
[3,]	"AR"	"32"	"0"	"1"	"2"
[4,]	"CA"	"209"	"1"	"0"	"1"
[5,]	"CO"	"56"	"1"	"1"	"2"
[6,]	"CT"	"20"	"0"	"0"	"0"
[7,]	"DE"	"5"	"0"	"0"	"0"
[8,]	"DC"	"4"	"0"	"0"	"0"
[9,]	"FL"	"201"	"1"	"1"	"4"
[10,]	"GA"	"51"	"1"	"0"	"1"

[11,]	"ID"	"8"	"0"	"0"	"0"
[12,]	"IL"	"92"	"1"	"0"	"1"
[13,]	"IN"	"38"	"0"	"0"	"0"
[14,]	"IA"	"47"	"0"	"1"	"2"
[15,]	"KS"	"17"	"0"	"0"	"0"
[16,]	"KY"	"30"	"1"	"0"	"1"
[17,]	"LA"	"70"	"1"	"1"	"2"
[18,]	"ME"	"23"	"0"	"1"	"2"
[19,]	"MD"	"33"	"0"	"0"	"0"
[20,]	"MA"	"42"	"0"	"0"	"0"
[21,]	"MI"	"150"	"0"	"1"	"2"
[22,]	"MN"	"82"	"0"	"1"	"2"
[23,]	"MS"	"18"	"0"	"0"	"0"
[24,]	"MO"	"92"	"1"	"1"	"2"
[25,]	"MT"	"6"	"0"	"0"	"0"
[26,]	"NE"	"11"	"0"	"0"	"0"
[27,]	"NV"	"22"	"1"	"1"	"2"
[28,]	"NH"	"20"	"0"	"1"	"2"
[29,]	"NJ"	"51"	"0"	"0"	"0"
[30,]	"NM"	"22"	"0"	"1"	"2"
[31,]	"NY"	"105"	"0"	"0"	"0"
[32,]	"NC"	"55"	"1"	"0"	"1"
[33,]	"ND"	"5"	"0"	"0"	"0"
[34,]	"OH"	"400"	"0"	"1"	"3"
[35,]	"OK"	"26"	"1"	"0"	"1"
[36,]	"OR"	"53"	"0"	"1"	"2"
[37,]	"PA"	"172"	"1"	"1"	"2"
[38,]	"RI"	"7"	"0"	"0"	"0"
[39,]	"SC"	"31"	"1"	"0"	"1"
[40,]	"SD"	"6"	"1"	"0"	"1"
[41,]	"TN"	"33"	"0"	"0"	"0"
[42,]	"TX"	"105"	"0"	"0"	"0"
[43,]	"UT"	"12"	"0"	"0"	"0"
[44,]	"VT"	"4"	"0"	"0"	"0"
[45,]	"VA"	"44"	"0"	"0"	"0"
[46,]	"WA"	"87"	"1"	"1"	"2"
[47,]	"WV"	"25"	"0"	"1"	"2"
[48,]	"WI"	"91"	"1"	"1"	"2"
[49,]	"WY"	"3"	"0"	"0"	"0"

```

> anhnps <- c(0, 0, 0, 0, 0)
> anhnps[1] <- sum(ninps[bg == 0])
> anhnps[2] <- sum(ninps[bg == 1])
> anhnps[3] <- sum(ninps[bg == 2])
> anhnps[4] <- sum(ninps[bg == 3])
> anhnps[5] <- sum(ninps[bg == 4])
> print(anhnps)

[1] 600 500 1101 400 201

> an <- sum(ninps)
> awhnps <- anhnps/((Npoph/sum(Npoph)) * an)
> nallsenatebg <- sum(ninps * bgsen)
> nallpresbg <- sum(ninps * bgpres)
> print(c(nallsenatebg, nallpresbg))

[1] 1291 1702

> senatepoph <- c(sum(to2000[bgsen == 1 & state !=
+ "FL"]), to2000[state == "FL"])
> senatenh <- c(sum(ninps[bgsen == 1 & state != "FL"]),
+ ninps[state == "FL"])
> senatepq <- c(0.25, 0.25)
> senatese <- sqrt(sum(((senatepoph/sum(senatepoph))^2) *
+ ((senatepoph - senatenh)/senatepoph) * senatepq/senatenh))
> senatemoe <- 2 * senatese
> print(c(senatepoph, senatenh, senatepq))

[1] 4.4563e+07 6.0060e+06 1.0900e+03 2.0100e+02 2.5000e-01
[6] 2.5000e-01

> print(c(senatese, senatemoe))

[1] 0.01398756 0.02797513

> prespoph <- c(sum(to2000[bgpres == 1 & state != "FL" &
+ state != "OH"]), to2000[state == "OH"], to2000[state ==
+ "FL"])
> presnh <- c(sum(ninps[bgpres == 1 & state != "FL" &
+ state != "OH"]), ninps[state == "OH"], ninps[state ==
+ "FL"])

```

```

> prespq <- c(0.25, 0.25, 0.25)
> presse <- sqrt(sum(((prespoph/sum(prespoph))^2) *
+ ((prespoph - presnh)/prespoph) * prespq/presnh))
> presmoe <- 2 * presse
> print(c(prespoph, presnh, prespq))

[1] 3.1918e+07 4.8230e+06 6.0060e+06 1.1010e+03 4.0000e+02
[6] 2.0100e+02 2.5000e-01 2.5000e-01 2.5000e-01

> print(c(presse, presmoe))

[1] 0.01261337 0.02522674

> nobgpop <- sum(to2000[bg == 0])
> nobgn <- sum(ninps[bg == 0])
> nobgpq <- 0.65 * 0.35
> nobgse <- sqrt(nobgpq/nobgn)
> nobgmoe <- 2 * nobgse
> print(c(nobgpop, nobgn, nobgpq))

[1] 4.0015e+07 6.0000e+02 2.2750e-01

> print(c(nobgse, nobgmoe))

[1] 0.01947220 0.03894440

> OHse <- sqrt(0.25/anhnp[4])
> OHmoe <- 2 * OHse
> FLse <- sqrt(0.25/anhnp[5])
> FLmoe <- 2 * FLse
> POPpoph <- c(sum(to2000[bg == 0]), sum(to2000[bg ==
+ 1]), sum(to2000[bg == 2]), to2000[state == "OH"],
+ to2000[state == "FL"])
> POPnh <- c(sum(ninps[bg == 0]), sum(ninps[bg == 1]),
+ sum(ninps[bg == 2]), ninps[state == "OH"], ninps[state ==
+ "FL"])
> POPpq <- c(0.65 * 0.35, 0.25, 0.25, 0.25, 0.25)
> POPse <- sqrt(sum(((POPpoph/sum(POPpoph))^2) * ((POPpoph -
+ POPnh)/POPpoph) * POPpq/POPnh))
> POPmoe <- 2 * POPse
> print(c(POPpoph, POPnh, POPpq))

```

```
[1] 4.0015e+07 2.7453e+07 3.1918e+07 4.8230e+06 6.0060e+06  
[6] 6.0000e+02 5.0000e+02 1.1010e+03 4.0000e+02 2.0100e+02  
[11] 2.2750e-01 2.5000e-01 2.5000e-01 2.5000e-01 2.5000e-01
```

```
> print(c(POPse, POPmoe))
```

```
[1] 0.01024369 0.02048737
```

## 2 Revised sample with new Florida weights

In Wave 1 Florida has 201 cases. This changes the weights and it slightly changes the standard errors for the Senate and Presidential battlegrounds and for the national estimates. None of these changes are substantial (except the standard error for Florida itself, of course.)

The table of weights is the practical import— Florida's weight in Wave 1 will reflect it's smaller sample size. Other strata weights are adjusted as well so that the weighted  $n$  will equal our actual sample size of 2802.

The sample sizes for each stratum are given in Table 1.

With this sample design, the standard errors and margins of error within each stratum and for the population are given in Table 2.

The sample can be weighted to provide population inferences while maintaining a weighted  $n$  equal to the actual  $n$  of 2802 (within rounding error!). These weights and the weighted  $n$  are given in Table 3.

The allocation of cases to states is given in Table 4.

Stratum	Sample size
NonBattleground	600
Senate only battleground	500
Presidential battleground	1101
Ohio	400
Florida	201
ALL Senate battleground	1291
Total	2802

Table 1: Sample sizes for each stratum, with oversampling of battleground states and state samples for Ohio and Florida. The sample size for ALL Senate battleground respondents includes those that are also in presidential battleground states and the 400 case oversample in Florida.

Stratum	Std. Error	Margin of Error
NonBattleground	0.0195	0.0389
All Senate battleground	0.014	0.028
Presidential battleground	0.0126	0.0252
Ohio	0.025	0.05
Florida	0.0353	0.0705
Population	0.0102	0.0205

Table 2: Standard errors and margin of error for each stratum. Note that the Senate here is for ALL Senate battleground cases, not those in the Senate ONLY battleground states. Estimates include the oversample cases from Ohio and Florida.

Stratum	Weight	Weighted $n$
Nonbattleground	1.696	1017
Senate only battleground	1.396	698
Presidential battleground	0.737	811
Ohio	0.307	123
Florida	0.76	153
ALL Senate battleground		1281

Table 3: Sample weights and weighted sample size for each stratum. The last line is the sample size for all Senate battleground respondents.

As a final check of the results, I use the calculated weights to weight our sample sizes as a proportion of the total sample. This is then compared to the state proportions of the population. The results should be a nearly perfect match, save for rounding of sample sizes to integers.

```
> pisample <- rep(0, 49)
> for (i in 0:4) {
+   pisample[bg == i] <- (ninps[bg == i]/awhmps[i +
+     1])/an
+ }
> pipop <- to2000/sum(to2000)
> summary(lm(pisample ~ pipop))
```

Call:

```
lm(formula = pisample ~ pipop)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.693e-04	-9.440e-05	-6.566e-06	7.816e-05	3.025e-04

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.264e-06	3.169e-05	0.071	0.943
pipop	9.999e-01	1.115e-03	896.504	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0001543 on 47 degrees of freedom  
Multiple R-Squared: 0.9999, Adjusted R-squared: 0.9999  
F-statistic: 8.037e+05 on 1 and 47 DF, p-value: < 2.2e-16

```
> plot(pisample ~ pipop)
> abline(0, 1)
```

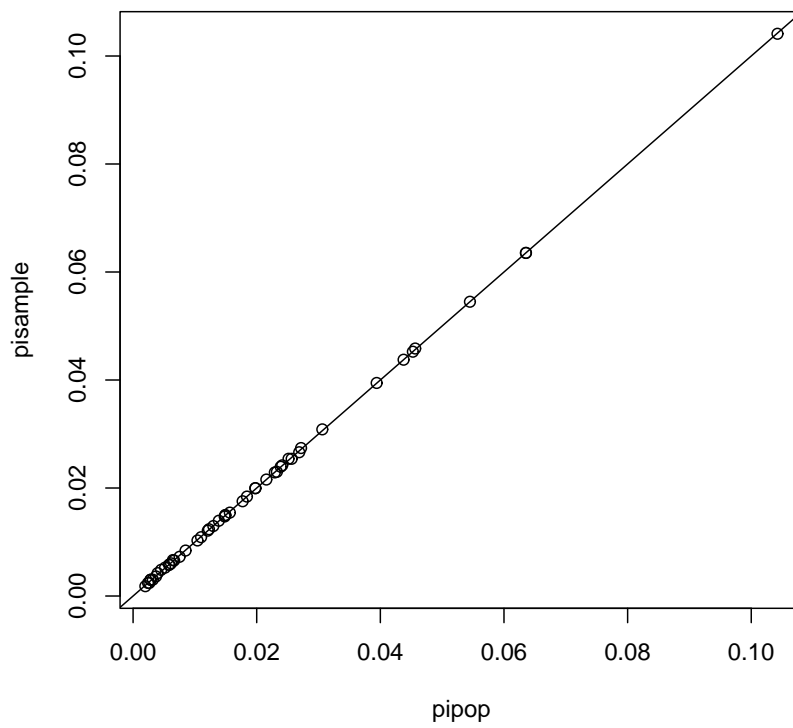


Figure 1: Weighted sample proportions vs population proportions

	State	n	State	n
1	AL	29	NE	11
2	AZ	57	NV	22
3	AR	32	NH	20
4	CA	209	NJ	51
5	CO	56	NM	22
6	CT	20	NY	105
7	DE	5	NC	55
8	DC	4	ND	5
9	FL	201	OH	400
10	GA	51	OK	26
11	ID	8	OR	53
12	IL	92	PA	172
13	IN	38	RI	7
14	IA	47	SC	31
15	KS	17	SD	6
16	KY	30	TN	33
17	LA	70	TX	105
18	ME	23	UT	12
19	MD	33	VT	4
20	MA	42	VA	44
21	MI	150	WA	87
22	MN	82	WV	25
23	MS	18	WI	91
24	MO	92	WY	3
25	MT	6		

Table 4: State sample sizes.